An Oriented SAR Ship Detector With Mixed Convolution Channel Attention Module and Geometric Nonmaximum Suppression

Chunnan Li and Haitao Lang, *Member, IEEE*

***Abstract*—Benefiting from deep learning, synthetic aperture radar (SAR) ship detection based on convolutional neural network (CNN) has developed rapidly and corresponding performance is getting better. Nevertheless, most of the existing methods still cannot achieve a good balance between *precision* and *recall* in scenes with complex background interferences, or in a scene where two or more ships dock side by side. To address these problems, this article proposes a novel oriented SAR ship detector, which uses oriented bounding boxes (OBBs) to describe ships. For the purpose of reducing missed ships (aiming to improve *recall*) while suppressing false alarms (aiming to maintain *precision*), the proposed detector embeds a mixed convolution channel attention (MCCA) module into the backbone network, which highlights the important feature channels to enhance ship representation features by reweighting all channels of the feature map. In addition, we consider the geometric position relationship of neighbor ships and propose geometric nonmaximum suppression (G-NMS) to remove the redundant ship candidates or possible false alarms. Extensive experiments conducted on the SSDD and HRSID$_s$ datasets demonstrate the effectiveness of MCCA and G-NMS, the proposed detector also achieves better performance compared to state-of-the-art OBB-based detectors.**

***Index Terms*—Geometric nonmaximum suppression (G-NMS), mixed convolution channel attention (MCCA), ship detection, synthetic aperture radar (SAR).**

I. Introduction

THANKS to its all-weather and all-day working mechanism, synthetic aperture radar (SAR) has wide applications, such as ship detection [1], [2], [3], [4], [5], [6], ship classification [7], [8], [9], oil spill detection [10], [11], SAR image segmentation [12], [13], [14], [15]. Among them, ship detection is a popular spot in remote sensing community, since it plays an important role in both military and civilian fields. Specifically, it is conducive to maritime traffic management, marine pollution control, combating illegal fishing, and defense and maritime security [16]. Nowadays, more and more efforts are devoted to ship detection task based on SAR image. Traditional SAR ship detection algorithm mainly relies on the statistical analysis of image pixels and establishes a model to identify ship pixels from sea clutter pixels [17], [18], [19], [20]. However, these methods often require many complex calculations, are very time-consuming, and have difficulties in achieving satisfactory results in complex backgrounds. In addition, the traditional method will lose a lot of ship texture information, making it difficult to meet the task demands.

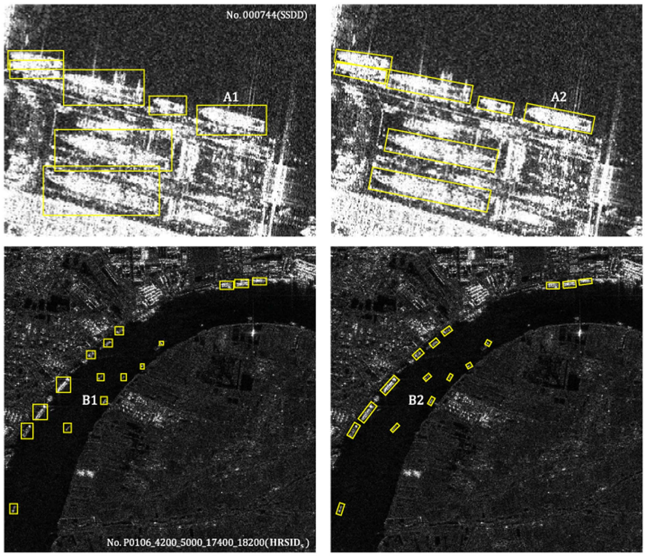
With the rapid development of deep learning, object detection based on convolutional neural network (CNN) has shined in the field of natural images [21], [22]. According to whether a candidate region proposal is used or not, CNN-based detectors are divided into two-stage and one-stage detectors. The role of the candidate region proposal is to generate rough areas where there may be targets, then the detector precisely finds targets in those areas. Representative two-stage detectors are Faster R-CNN [23], mask R-CNN [24], cascade R-CNN [25]. The one-stage detector is a direct regression to predict the position and category of a target, thus improving the detection speed. Representative one-stage detectors are SDD [26], YOLOv1-v4 [27], [28], [29], [30]. The success of CNN-based object detection method makes it widely used in image object detection tasks [31], [32], [33], it has also enlightened SAR ship detection research. Although the imaging mechanism is different between natural images and SAR images, it does not prevent the usage of CNN-based object detection methods for SAR ship detection.

In recent years, a large number of studies on CNN-based SAR ship detection have emerged. Chang et al. [34] used YOLOv2 to achieve rapid SAR ship detection. To solve multiscale ship detection, Cui et al. [35] proposed a dense attention pyramid network, Deng et al. [36] used dense connection operations to achieve feature reuse and used filters with different sizes in the candidate region proposal stage to produce as many ship regions at different scales as possible, thereby improving recall. Since the complex background interferences in SAR images can easily cause false alarms, some methods add attention modules to the feature extraction network to highlight important ship features. Lin et al. [37] added squeeze and excitation (SE) to Faster R-CNN. Zhao et al. [38] used dilated convolution and convolutional block attention module to refine ship features. Yang et al. [39] designed a coordinate attention module to regress the more accurate position of ships. Besides, considering that optical remote sensing images are similar to SAR images (they are both bird's-eye views) and ships in optical remote sensing images usually have rich texture features, Bao et al. [40] pretrained

Manuscript received 27 April 2022; revised 28 May 2022 and 11 August 2022; accepted 9 September 2022. Date of publication 13 September 2022; date of current version 23 September 2022. This work was supported by the National Natural Science Foundation of China under Grant 62071030. *(Corresponding author: Haitao Lang.)*

The authors are with the College of Mathematics and Physics, Beijing University of Chemical Technology, Beijing 100029, China (e-mail: chunnan@buct.edu.cn; haitaolang@hotmail.com).

Digital Object Identifier 10.1109/JSTARS.2022.3206247



(a) (b)

Fig. 1. Different ship expressions. (a) HBB. (b) OBB. Each yellow rectangle boxes a ship.

CNN with optical remote sensing ship images and used bridge neural network to learn common features of ships in optical and SAR images to extract SAR ship texture features. Furthermore, to address the problem of low ship detection performance in inshore scenes, Zhang et al. [41] improved the detection performance by increasing the number of inshore samples in the training process. Yang et al. [42] adopted to assign different loss weights to inshore samples and offshore samples to suppress false alarms.

However, most existing CNN-based SAR ship detection methods apply a horizontal bounding box (HBB) to describe a ship, such as abovementioned methods. The main disadvantages of HBB are that it contains not only ship, but also many useless backgrounds, and ships in SAR images are in arbitrary orientations so it is difficult to get the geometric information of a ship from its HBB. Recent research works proposed using an oriented bounding box (OBB) to describe a ship. OBB is essentially a rectangle closest to a target, which can be rotated arbitrarily according to the direction of a ship. As shown in Fig. 1, A1 and A2 box the same ship but with HBB and OBB respectively. A1 contains not only the ship area but also some land areas and sea areas, but most of it is the ship area and only a small land area or sea area in A2. In addition, HBB cannot keep the aspect ratio information of ship better, especially when ships appear small in SAR images. For example, B1 has lost the aspect ratio information of the ship, while B2 keeps the aspect ratio of the ship well. Yang et al. [43] proposed an OBB-based ship detector, which uses adaptive intersection over union (IoU) threshold to assign positive samples and selects suitable feature maps based on the scale of ships to avoid feature redundancy. Sun et al. [44] added an angle classification structure to the detection head to obtain the angle information of ships. He et al. [45] used polar coordinates to encode OBB to solve the boundary discontinuity problem. Anchor is a common means in CNN-based object detection methods, which helps to improve recall and is widely used in both HBB and OBB methods. Anchor-based methods usually require precise presetting of the size and aspect ratio of anchors and a lot of calculations. Therefore, some anchor-free SAR ship detection methods are proposed. Cui et al. [46] used spatial shuffle-Group enhance attention to aggregate more ship features between channels, detected the keypoints of a ship to get its HBB, and achieved large-scale SAR image ship detection. Fu et al. [47] used an attention-guided balanced pyramid that adaptively learns the weights of different-level feature maps for fusion, and refines the feature maps to enhance the focus on the ship region. Sun et al. [48] designed a category-position module to optimize ship position regression, which can improve ship localization performance by generating guidance vectors from classification branch features. Zhang et al. [49] proposed an oriented anchor-free detector by using an oriented non-normalized Gaussian function to guide the regression of ship rotation angle. Fu et al. [50] guided the network for ship feature learning by clustering the scattering keypoints on ships, while an angle prediction branch was used to obtain the rotation angle of a ship.

As far as SAR ship detection task is concerned, CNN-based methods have been developed in many varieties with different advantages. As presented above, many difficulties have been solved, such as ship feature extraction and fusion, multi-scale ship detection, and large-scene ship detection. Nevertheless, there are still many issues that have not been solved well. Firstly, limited by the distinguishing ability of ship representation features, real ships are often misidentified as land backgrounds (missed ships), resulting in poor ship detection *recall*, and the opposite situation also occurs (false alarms). Secondly, it is necessary to use a post-processing algorithm to remove the redundant detection results (usually using non-maximum suppression (NMS) [51]), but original NMS usually cannot effectively remove all repeated detections leading to a decrease in *precision*. In OBB-based methods, it also tends to remove high-quality detection candidates but retain low-quality counterparts produced due to poor angle regression, especially in the scene where two or more ships dock side by side.

According to the previous analysis, we believe that CNN is suitable for SAR ship feature extraction, and compared to HBB, OBB contains fewer background interferences and is more suitable as the bounding box (bbox) of a ship. Therefore, we propose a novel OBB-based SAR ship detector to handle the aforementioned issues.

Our main contribution is threefold.

1) We propose a mixed convolution channel attention (MCCA) module to reassign new weight to each channel of a feature map by considering both the relationship between neighbor channels and the relationship between all channels of the feature map. MCCA is embedded into the backbone network to highlight important channels and weaken useless channels of feature maps to extract more discriminative ship features, which effectively improve *recall* by reducing missed ships while maintaining *precision* by not increasing false alarms.

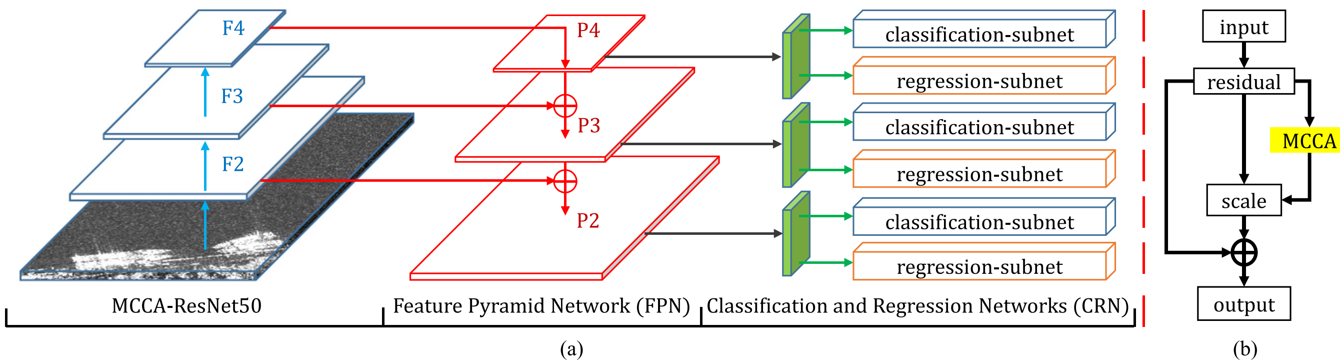


Fig. 2. The core structure of the proposed detector. (a) Main workflow and three components. MCCA-ResNet50 is used for ship feature extraction, which sequentially extracts shallow and deep features (bottom-top, the resolution of the feature map gradually decreases, but the semantic information is richer.). FPN is used to fuse shallow and deep features (top-bottom, the resolution of the deep feature map is increased by upsampling and fused with the shallow feature map at the same resolution). During the CRN, feature maps at three scales are selected for ship detection to ensure that both small and large ships are accurately detected. (b) MCCA-ResNet module, the basic module in MCCA-ResNet50, is used to reweight each channel of feature maps. ⊕ means element-wise addition.

2) We innovatively propose G-NMS after analyzing the issue that arises in OBB-based ship detection and the geometric position relationship of neighbor ships in real situations. G-NMS aims to remove the low-quality and/or redundant detection results more accurately by calculating the center point distances between different ship candidates, which further improves *precision*.
3) We evaluate the proposed methods on SSDD [52] and HRSID$_s$[1] [53] datasets to verify their improvements. Compared with several state-of-the-art detectors, the proposed detector achieves a good balance between *recall* and *precision*, and outperforms comparison detectors by 1%–2% in terms of average precision (AP).

In the subsequent content of this article, Section II details our proposed MCCA and G-NMS, and Section III conducts ablation experiments and compares with other state-of-the-art detectors on two datasets. Finally, Section IV concludes this article.

II. Proposed Methods

A. Overall Workflow

We propose a one-stage OBB-based ship detector, which is established based on R^{3}Det [54]. The framework of the proposed detector is shown in Fig. 2(a), which has three successive components: backbone network for feature extraction, feature pyramid network (FPN) for feature fusion, and classification and regression networks (CRN) for ship classification and location. We design MCCA-ResNet50 as the backbone network and select the feature maps at the last three scales, i.e., F2, F3, F4 whose size is 1/8, 1/16, 1/32 of the input image. The larger the size of the feature map, the more accurate the location information, while the smaller the size of the feature map, the richer the semantic information. In order to adapt to ship detection at different scales, we use FPN to fuse F2, F3, F4 to get P2, P3, P4. In this way, feature maps at three scales can contain abundant semantic and location information. In the CRN step (consistent with R^{3}Det [54]), the classification subnet is used to determine whether a detected object is a ship, and the regression subnet is used to locate the position information of a confirmed object (the OBB of an object is defined by five parameters (x, y, w, h, θ), where (x, y), w, h denote the center point, width, height, $\theta \in[-90^\circ, 0]$ is its rotation angle.), then subnets map classification and location results at all scales back to their original image. After that, we use G-NMS to remove possible false alarms and redundant detection outputs (i.e., single ship marked by multiple OBBs).

B. MCCA-ResNet50

In SAR image, since many land buildings have similar brightness and shape to ships, these buildings are often misidentified as ships or vice versa, leading to false alarms or missed ships. To handle this problem, it is necessary to highlight important parts in ship feature maps, for which we design MCCA and embed it into the bottleneck of ResNet50 [see Fig. 2(b)] to extract more discriminative ship representation features.

MCCA is a channel attention module, which aims to reweight each channel of a feature map so that important channels get a higher weight, and useless channels get a lower weight, thereby enhancing the representation ability of the feature map. In order to comprehensively consider the importance of each channel of a feature map, MCCA applies alternately two-dimensional convolution operation (conv2d, kernel size is 1×1) to establish the relationship between a channel and all channels of the ship feature map, one-dimensional convolution (conv1d) to establish the relationship between neighbor channels, and its process of reassigning channel weights is shown in Fig. 3(a). The difference between conv2d and conv1d is shown in Fig. 3(b), and the kernel number k of conv1d is set to 5. Compared with other channel attention modules SE [55] and efficient channel attention (ECA) [56], SE reweights each channel of the feature map by fully-connected (FC) operation, so the weight of each channel is only the result of considering all channels [see Fig. 3(c)]. Contrary to SE, ECA only focuses on the relationship between neighbor channels of the feature map, so it uses conv1d to reassign new weight to each channel of the feature map, as

[1]This article uses a portion of HRSID, hence names it as HRSID$_s$.

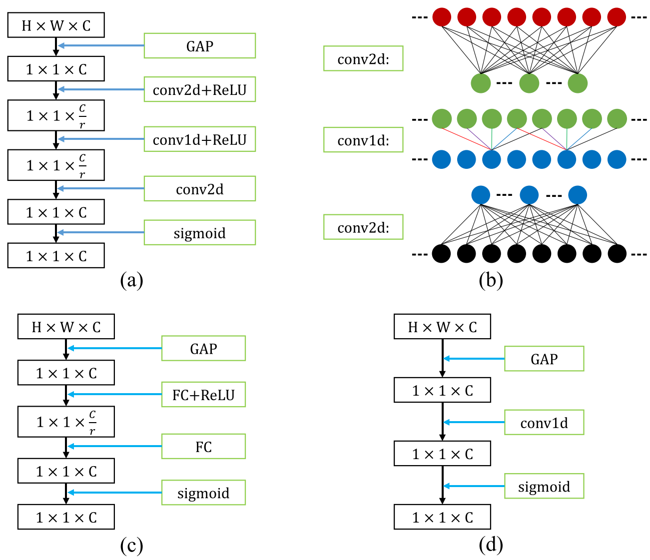


Fig. 3. Different attention modules. (a) MCCA. (b) The difference between conv1d and conv2d operation. (c) SE. (d) ECA. H, W, C are the height, width and channel number of the input feature map, r is channel number attenuation factor. GAP is global average pooling, FC is fully connected, ReLU and sigmoid are activation function.

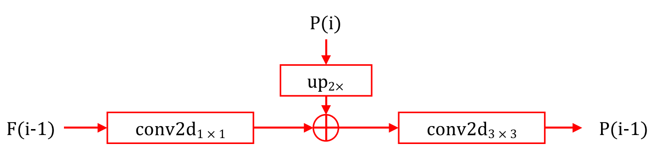


Fig. 4. Detail in FPN. 1×1 and 3×3 represent the kernel size of conv2d.

shown in Fig. 3(d). We believe that MCCA jointly considers both all channels and neighbor channels, the reweighted weights will be more helpful to enhance the distinguishing ability of representation features.

C. Feature Fusion

Ships in SAR images are different scales, so the feature fusion structure is used to accommodate ship detection at different scales. The shallow features extracted by the backbone network are rich in position information and the deep features are rich in semantic information. Therefore, the deep and shallow features are fused to obtain more informative ship features, as shown in Fig. 4. The $\mathrm{up}_{2\times}$ means $2\times$ upsampling, its purpose is to increase the scale of P(i) to be consistent with F(i-1), $\mathrm{conv2d}_{1\times1}$ is used to adjust channel numbers of F(i-1) to be consistent with P(i), $\oplus$ means element-wise addition, $\mathrm{conv2d}_{3\times3}$ is used to refine the feature after element-wise addition. Notably, F4 is the deepest feature, so a conv2d with kernel size 3×3 is directly applied to F4 to obtain P4. The process is formulated as follows:

$$\mathrm{P4} = \mathrm{conv2d}_{3\times3}(\mathrm{F4}) \tag{1}$$

$$\begin{aligned}\mathrm{P(i-1)} &= \mathrm{conv2d}_{3\times3}(\mathrm{conv2d}_{1\times1}(\mathrm{F(i-1)}) \\ &\quad + \mathrm{up}_{2\times}(\mathrm{P(i)})), \mathrm{i} = 3, 4.\end{aligned} \tag{2}$$

D. G-NMS

CNN-based detectors usually use NMS to remove repeated candidates from the detection results, which is expressed as

$$c_i = \begin{cases} c_i, & \mathrm{IoU}(b_i, b_{\max}) \le \mathrm{T} \\ 0, & \mathrm{IoU}(b_i, b_{\max}) > \mathrm{T} \end{cases} \tag{3}$$

where b_i is a bbox whose classification confidence is c_i, and $b_{\max}$ is the bbox with maximum confidence. It shows that NMS selects the one with high confidence as the final result for two or more intersecting bboxes with high overlap, while for those with a low overlap (i.e., IoU $\le$ T), it keeps them directly. For the OBB method, the angle difference between a predicted bbox and the corresponding ground truth (GT) largely determines the quality of the predicted box. In general, the larger this difference, the lower the quality of the predicted bbox. The original NMS is difficult to effectively balance how to preserve high-quality OBB with low confidence without increasing false alarms. As shown in Fig. 5, it is more expected to retain high-quality OBBs (green) and remove low-quality ones (red). However, for original NMS, when T is set to a small value, some high-quality OBBs with low confidence are removed [see Fig. 5(b)], while when T is set to a large value, although high-quality OBBs are preserved, a large number of low-quality OBBs are also retained [see Fig. 5(c)].

To solve the above problem, we carefully consider the geometric position relationship of neighbor ships and conclude that the distance between the center points of two ships must be higher than a specific threshold, no matter how close they are. With this idea, we propose G-NMS that considers the geometric position relationship to handle intersecting bboxes. First, assuming that $\mathrm{Q} = \{b_1, b_2, \ldots, b_N\}$ is the set of detection bboxes, and the corresponding confidences are $\{c_1, c_2, \ldots, c_N\}$ which are sorted from smallest to largest, i.e., $c_1 \le c_2 \le \ldots \le c_N$. Next, we select b_i $(i = 1, \ldots, N)$ from Q one by one in order, and compare it with other bboxes b_j $(j = i + 1, \ldots, N)$ with greater confidences according to the following:

$$c_i = \begin{cases} 0, & \mathrm{IoU}(b_i, b_j) > \mathrm{T} \\ \alpha * c_i, & \mathrm{IoU}(b_i, b_j) \le \mathrm{T}, \rho \le m, c_j - c_i \le n \\ \beta * c_i, & \mathrm{IoU}(b_i, b_j) \le \mathrm{T}, \rho \le m, c_j - c_i > n \\ c_i, & \mathrm{IoU}(b_i, b_j) \le \mathrm{T}, \rho > m \end{cases} \tag{4}$$

where ρ is the Euclidean distance between the center points of b_i and b_j whose sizes are $w_i \times h_i$ and $w_j \times h_j$, respectively. $m = \min\{w_i, h_i, w_j, h_j\}$, n is the threshold of the difference in confidences. $\alpha, \beta \in (0, 1)$ are attenuation factors. When IoU(b_i, b_j) is higher than T, since $c_i \le c_j$, the confidence of b_i is set to zero. When IoU(b_i, b_j) is less than T, the following conditions are considered, respectively: 1) If b_i, b_j are very close ($\rho \le m$) and have similar confidences ($c_j - c_i \le n$), they are most likely to describe the same ship and a large attenuation factor α is used to attenuate c_i. 2) If b_i, b_j are very close ($\rho \le m$) but have a large difference ($c_j - c_i > n$) in confidence, they are more likely to describe two different ships and a small attenuation factor β is used to attenuate c_i. 3) If b_i, b_j are far apart, $\rho > m$, their confidences are not adjusted. During this process, c_i is at most attenuated once. Through the above operations, the confidence of low-quality detection results will be attenuated to a very low

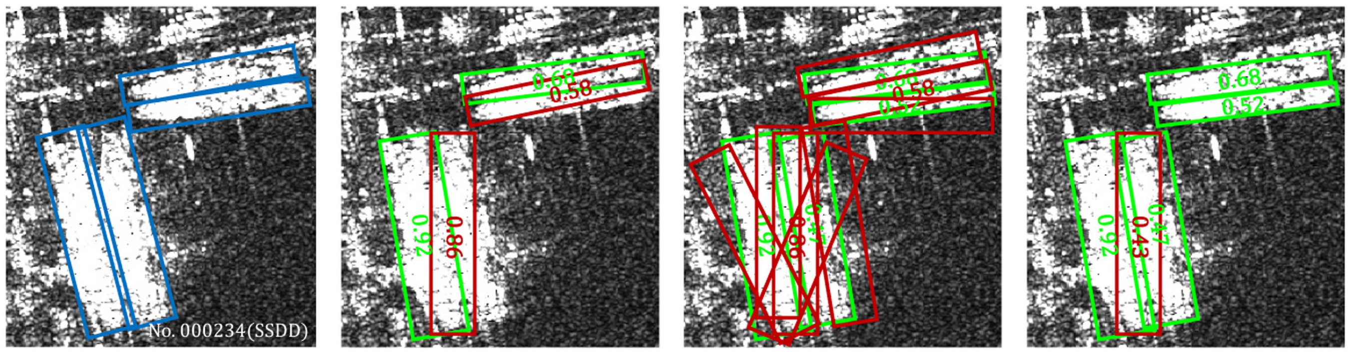


Fig. 5. Redundant candidate elimination. (a) GT. (b) NMS with small T. (c) NMS with large T. (d) G-NMS with large T. The blue, green, and red represent GT, high-quality and low-quality OBBs.

level. In practical applications, low-quality or repeated results can be removed by setting the relevant condition thresholds (i.e., $\mathrm{T}, n, \alpha, \beta$) reasonably, and high-quality results can be retained. An example is shown in Fig. 5(d).

E. Loss Function

The loss function consists of classification loss (L_{cls}) and regression loss (L_{reg}). Classic focal loss (FL) [57] is used as L_{cls}, which is formulated as

$$L_{\text{cls}} = \text{FL} = -\alpha_t(1-p_t)^{\gamma} log(p_t) \tag{5}$$

where α_t is weighting factor, γ is tunable focusing parameter, $p_t \in$(0,1] is the model's estimated probability.

In this article, a predicted OBB of an anchor box is $B = (x_p, y_p, w_p, h_p, \theta_p)$, and the corresponding GT of the anchor box is $G = (x_g, y_g, w_g, h_g, \theta_g)$. To avoid the loss discontinuity caused by angle periodicity and sudden change of related width and height, L_{reg} is set based on [58].

$$L_c = s(|x_p - x_g|) + s(|y_p - y_g|) \tag{6}$$

$$L_{r1} = s(|w_p - w_g|) + s(|h_p - h_g|) + s(|\theta_p - \theta_g|) \tag{7}$$

$$L_{r2} = s(|w_p - h_g|) + s(|h_p - w_g|) + s(|90 - |\theta_p - \theta_g||) \tag{8}$$

$$L_r = L_c + \min\{L_{r1}, L_{r2}\} \tag{9}$$

$$L_{\text{reg}} = \sum_{i \in N_p} L_r^i. \tag{10}$$

N_p represents the number of positive samples, $s(\cdot)$ represents smooth-L1 [59].

Thus, the total loss function is expressed as

$$L_{\text{total}} = \frac{1}{N_p} L_{\text{cls}} + \frac{1}{N_p} L_{\text{reg}} \tag{11}$$

III. Experiments and Analysis

A. Datasets and Experimental Settings

1) Datasets: We employ SSDD [52] and HRSID$_s$ datasets to validate the proposed method. SSDD contains 1160 images (2456 ships), which are collected from various satellites with resolutions differing from 1 to 15 m. HRSID$_s$ is composed of 700 images (3546 ships) from randomly selected inshore and dense scenes in HRSID [53], it contains SAR images with a resolution of 0.5, 1, 3 m. Besides, there are a large number of small ships in HRSID$_s$ (as shown in the second row of Fig. 1), so detecting ships on HRSID$_s$ is more difficult than SSDD. Based on their original annotations (i.e., GT), we relabel the GT in OBB format to evaluate all comparison methods. In our experiments, we divide training set and test set into the ratio of 8:2, the test data of SSDD is the data whose file names end with 0 and 9, and test data of HRSID$_s$ is randomly selected. During the training, all images are resized to 800 $\times$ 800 and random flip is used to enhance datasets.

TABLE I
Parameters Setting

Parameter	r	k	n	T	α	β	α_t	γ
Our Setting	8	5	0.05	0.3	0.1	0.5	0.25	2

2) Parameters Setting: We set the aspect ratio of anchor as {1/5, 1/3, 1/2, 1, 2, 3, 5}, the rotation angle of anchor as $\{-90°, -75°, -60°, -45°, -30°, -15°\}$. The setting of other relevant parameters used in experiments is listed in Table I.

3) Implementation Details: All our experiments are based on MMDetection toolbox [60]. We use stochastic gradient descent (SGD) optimizer to train model, the initial learning rate is 0.005, momentum is 0.9, and weight decay is 0.0001. The experiments run on a computer with Intel i9-9980XE CPU 3.00 GHz and GeForce GTX 2080 Ti, the operating system is Ubuntu 20.04.3.

B. Evaluation Criteria

We evaluate detectors based on PR-curve, which shows the relationship between *precision* (P) and *recall* (R) for every possible cut-off (i.e., detection threshold). A detector with a higher AP, which is corresponding to a larger area under the curve (AUC) and has a better overall performance. It is desired that a detector should have both high P and high R. However, most detection algorithms often involve a trade-off between the two. Break-even point (BEP) is the point on PR-curve where

TABLE II
PERFORMANCE OF PROPOSED COMPOSITIONS ON TWO DATASETS

R^3Det + different moudle	SSDD		HRSID_s	
	AP(%)	BEP	AP(%)	BEP
R^3Det + NMS	93.86	0.930	82.73	0.827
R^3Det + NMS + MCCA	95.32	0.945	83.66	0.826
R^3Det + G-NMS	94.87	0.932	83.37	0.841
R^3Det + MCCA + G-NMS	96.47	0.949	84.32	0.840

TABLE III
PERFORMANCE OF DIFFERENT ATTENTION MECHANISMS

different attention	SSDD		HRSID_s	
	AP(%)	BEP	AP(%)	BEP
R^3Det + NMS + SE	94.98	0.942	82.84	0.828
R^3Det + NMS + ECA	94.80	0.934	83.01	0.835
R^3Det + NMS + MCCA	95.32	0.945	83.66	0.826

P = R, which is utilized to evaluate the balance performance. In this article, if the IoU of a predicted bbox with its corresponding GT is larger than 0.5, this predicted bbox is considered to be a correct detection

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{12}$$

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{13}$$

$$\mathrm{AP} = \int_0^1 \mathrm{P}(\mathrm{R})\mathrm{dR} \tag{14}$$

where TP, FP, FN are the number of correctly detected ships, false alarms, and missed ships, respectively.

C. Ablation Experiments

We take R^3Det [54] as the baseline and perform an ablation study on SSDD and HRSID_s, the experimental results are shown in Table II . First, we only add MCCA to R^3Det (with NMS). On SSDD dataset, AP and BEP are, respectively, improved by 1.46% and 0.015 compared to the baseline, which shows that after MCCA is added, more ships are detected, and accuracy is also higher. On HRSID_s dataset, AP increased by 0.93%, but BEP decreased by 0.001, which means that while improving *recall*, it also increases false alarms, but this is acceptable because there are a large number of small ships in HRSID_s dataset. Small ships in SAR images have little or no texture information, which leads to small ship features are very similar to some background features. When *recall* is improved, some backgrounds are also recognized as ships, resulting in a decrease in the accuracy. Next, when adding only G-NMS to R^3 Det. The improvement of AP on two datasets is not high, it is 1.01% on SSDD, and only 0.64% on HRSID_s. BEP is improved by 0.002 (slightly not obvious) on SSDD and 0.014 on HRSID_s, this reveals that after using G-NMS, false alarms are effectively suppressed, and *precision* is significantly improved, thereby improving BEP. Finally, we add both MCCA and G-NMS to R^3 Det, which achieves significant performance improvements on two datasets, AP and BEP improved by 2.61% and 0.019, respectively, on SSDD, 1.59% and 0.013 on HRSID_s, the results fully show that both *recall* and *precision* have been improved. In addition, it is worth noting that the possible false alarm of the detector on HRSID_s dataset due to adding MCCA (when only adding MCCA, BEP drops compared to the baseline) is well resolved after using G-NMS (while AP is improved, BEP is also significantly improved), which indicates that MCCA and G-NMS can complement each other and work together to improve detection performance.

D. Comparing the Proposed Compositions With Other Methods

1) Comparing MCCA With Other Attention Mechanisms: In order to validate the effectiveness of the proposed MCCA, we compare it with the SE and ECA attention mechanisms by adding three modules in turn to the R^3Det (with NMS). The detection results on two datasets are shown in Table III.

It can be seen that adding the attention module can improve AP of the detector, among which MCCA is better than SE and ECA. This result confirms our assumption that the reweighted weights obtained by jointly considering both all channels and neighbor channels are more helpful to enhance the distinguishing ability of representation features. However, it can be found that MCCA is lower than SE and ECA in terms of BEP on HRSID_s. This is because small ships are easily missed in detection process. Especially in inshore, improving *recall* often results in a decrease in *precision*. Fortunately, MCCA does not cause too much decrease in BEP (only 0.009 lower than ECA, and 0.002 lower than SE) while bringing considerable AP improvement.

2) Comparing G-NMS With Other Postprocessing Algorithms: In order to validate the effectiveness of the proposed G-NMS. Taking R^3Det as the basis, we replace NMS with Soft NMS [61], DIoU NMS [62], Weighted Boxes Fusion (WBF) [63], and G-NMS in turn to test the detection performance. As described in Section II-D, NMS takes a barbarous approach, i.e., it keeps the bbox with the highest confidence while zeroing the confidences of other bboxes with IoU larger than the threshold (it is equivalent to removing the bboxes). However, since the threshold is difficult to determine, NMS can easily cause missed detections or false alarms. To solve this problem, Soft NMS designs a confidence attenuation function based on IoU (the one with lower confidence is attenuated, and the larger the IoU, the more severe the confidence attenuation), which is different from NMS that directly sets confidence to 0. However, only using IoU to measure the distance between bboxes is not enough in dense scenes, so DIoU is proposed, which considers both IoU and center point distance between bboxes. DIoU NMS is formed by replacing the IoU in NMS with DIoU, and still retains the operation of directly attenuating confidence to 0. WBF abandons the idea of directly removing a bbox based on a certain measure (IoU or DIoU), and it performs a weighted fusion for bboxes whose IoU is larger than threshold to get a new bbox. In this process, the larger the confidence of a bbox, the greater the role it plays. However, the new bbox may be affected by the low-quality bboxes, resulting in poor fusion.

Fig. 6 shows the results of two samples (a and b) with different postprocessing algorithms. From (a-2), it can be seen that NMS

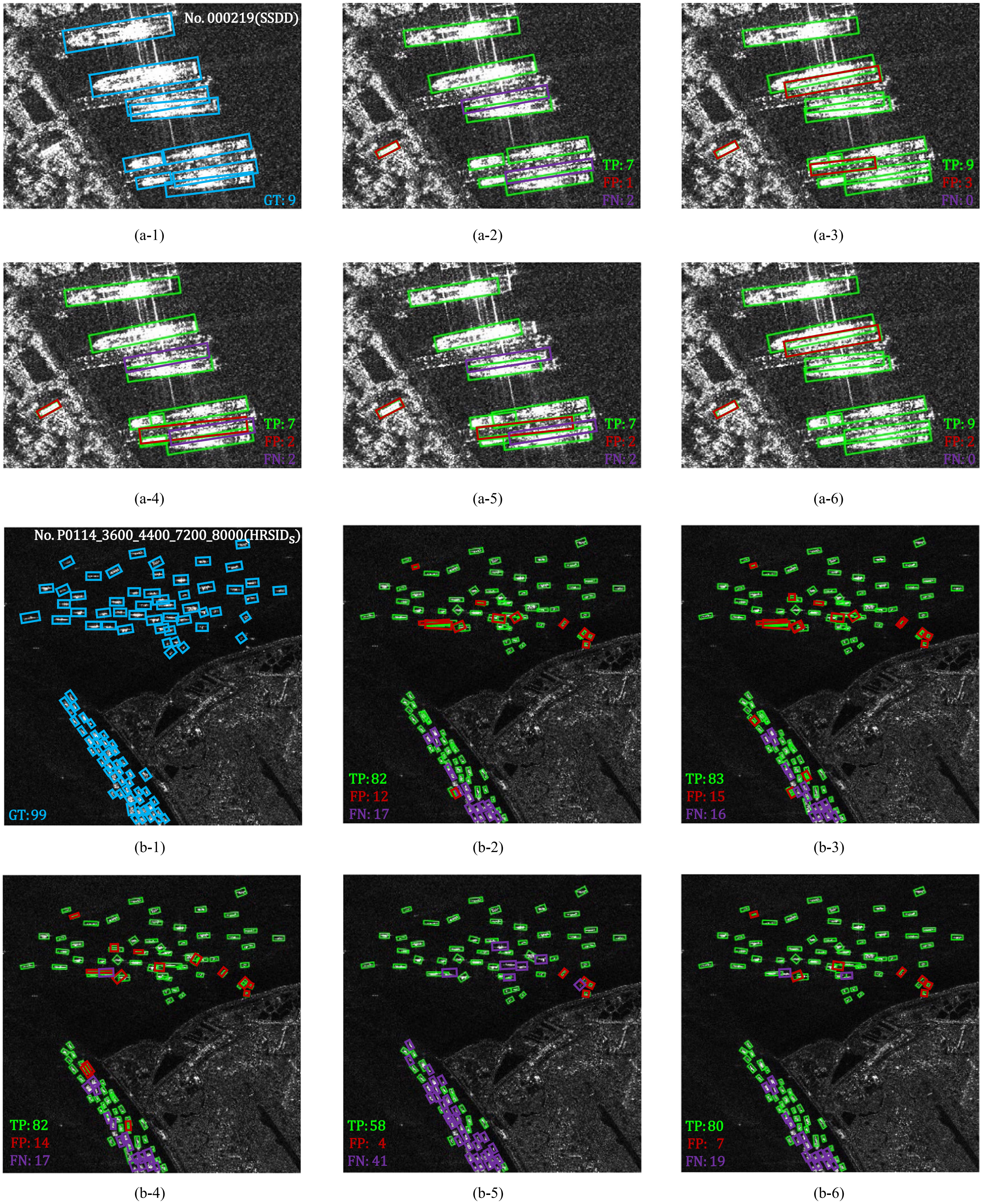


Fig. 6. Results of different post-processing algorithms. (a-1) and (b-1) are GT. (a-2) and (b-2) are results of NMS. (a-3) and (b-3) are results of Soft NMS. (a-4) and (b-4) are results of DIoU NMS. (a-5) and (b-5) are results of WBF. (a-6) and (b-6) are results of G-NMS. The blue, green, red, and purple rectangles represent respectively GT, correct detection (TP), false alarms (FP), and missed ships (FN).

TABLE IV
PERFORMANCE OF DIFFERENT POSTPROCESSING ALGORITHMS

different post-processing	SSDD		$HRSID_s$	
	AP(%)	BEP	AP(%)	BEP
R^3Det + NMS	93.86	0.930	82.73	0.827
R^3Det + Soft NMS [61]	94.18	0.932	82.90	0.832
R^3Det + DIoU NMS [62]	94.23	0.930	83.02	0.838
R^3Det + WBF [63]	93.15	0.921	76.51	0.822
R^3Det + G-NMS	94.87	0.932	83.37	0.841

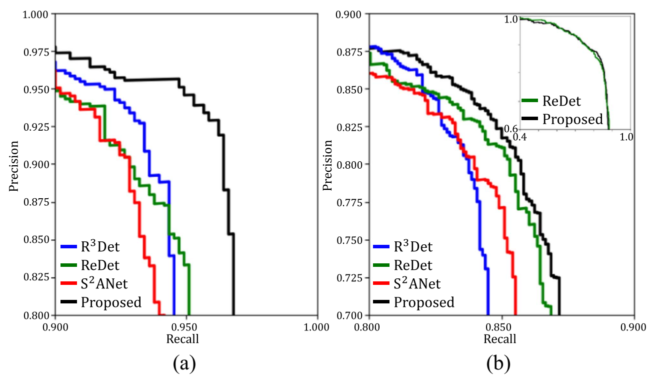


Fig. 7. PR-curves of different detectors on two datasets. (a) Results of different detectors on SSDD. (b) Results of different detectors on $HRSID_s$.

is prone to remove correct ship detections when ships are parked side by side. Soft NMS can reduce this situation, but it retains more false alarms. Compared with NMS, DIoU NMS leaves more false alarms with the same number of removing correct results, perhaps DIoU is approximately equal to IoU in some situations and DIoU NMS uses a larger threshold to cause it. Besides, comparing (a-2) and (a-5), it can be found that since the fusion of multiple OBBs causes high-quality OBBs to be corrupted, some OBBs do not surround the ships well. Therefore, WBF may not be suitable for the postprocessing of oriented SAR ship detection. Observing (a-6), G-NMS not only preserves high-quality detection results for all ships, but also leaves fewer false alarms than other methods. However, from the results of (b-2)–(b-6), it can be seen that the detection of small ships in dense scenes is very challenging and prone to false alarms and missed detections. The detector alone uses NMS, Soft NMS, DIoU NMS, and G-NMS are similar in missed detection, but G-NMS has the best effect of removing false alarms. Besides, WBF leaves the fewest false alarms among several methods, but it misses a large number of ships, which may be due to the low-quality fusion effect that causes some correct detection results to be removed. Although G-NMS also may remove some correct detections and retain some false alarms, it achieves the best results and outperforms comparative methods.

Next, we quantitatively analyze these postprocessing algorithms, the results are shown in Table IV. Compared to NMS, Soft NMS achieves better performance on two datasets, although the improvement is slightly small (mainly 0.32% AP improvement on SSDD, and 0.005 BEP improvement on $HRSID_s$), but this result still shows that it is not appropriate to directly attenuate the confidence of a bbox to 0. Soft NMS attenuates confidence of a bbox according to IoU, the larger the IoU, the more severe the confidence attenuation. This is because the larger the IoU, the greater the possibility that the attenuated bbox is redundant, so that high-quality nonredundant bboxes may be attenuated, but not removed directly. For DIoU NMS, the BEP on SSDD is consistent with NMS, and the AP is improved by 0.37%, but the AP is improved by 0.29% and the BEP is improved by 0.011 on $HRSID_s$. This result shows that the center points distance between bboxes is a factor worth considering because no matter how close two targets are, their center points are always different. Because of this, DIoU NMS achieves better performance than NMS. In addition, comparing soft NMS and DIoU NMS, it can be found that the performance of DIoU NMS is better than soft NMS, which once again illustrates the importance of the center points distance between bboxes. Meanwhile, the superior performance of these two algorithms over NMS proves that our idea of designing G-NMS is reasonable. However, what is surprising and expected is that when WBF is used instead of NMS, the performance of the detector drops on two datasets and AP on $HRSID_s$ is only 76.51%, which is 6.22% lower than NMS. The possible reason is that WBF adopts a coordinate weighted fusion method for the OBBs whose IoU are higher than the threshold. In the OBB method, the weighted fusion result of multiple OBBs is a quadrilateral instead of a rectangle, so the final result can only be taken as the smallest bounding rectangle of the resulting quadrilateral. During this process, many high-quality OBBs are destroyed, resulting in inaccurate position prediction [as shown in Fig. 6(a-5)] and causing a large number of ships to be missed (*recall* drops). Although *recall* is severely affected, the high *precision* still maintains the BEP at a stable level (0.009 and 0.005 lower than NMS on SSDD and $HRSID_s$, respectively). However, this is not what we want, we would like to achieve a higher level of both *recall* and *precision*.

Going back to the proposed G-NMS, it achieves the best performance on SSDD and $HRSID_s$ (AP and BEP are, respectively, 1.01% and 0.002 higher than NMS on SSDD, AP and BEP are, respectively, 0.64% and 0.014 higher than NMS on $HRSID_s$), while outperforming Soft NMS, DIoU NMS, and WBF. However, we can find that it is mainly the improvement of AP on SSDD and the improvement of BEP on $HRSID_s$. This is because SSDD contains a small number of small ships, mostly larger ships (larger ships are relatively easy to detect), it is easy to achieve a good balance between *recall* and *precision* so the difference between the BEP of different algorithms is not obvious, and the improvement of AP also means that the correct detections of more ships are retained after using G-NMS. Different from SSDD, there are a large number of dense small ships on $HRSID_s$, which is extremely challenging for the postprocessing algorithm. Although the improvement of AP by G-NMS is not very large, mainly because small ships are easier to be missed, the BEP gets an obvious improvement, which shows that G-NMS is beneficial not only for improving *recall*, but also for *precision*. In other words, these results fully demonstrate the effectiveness of G-NMS in dense ship detection.

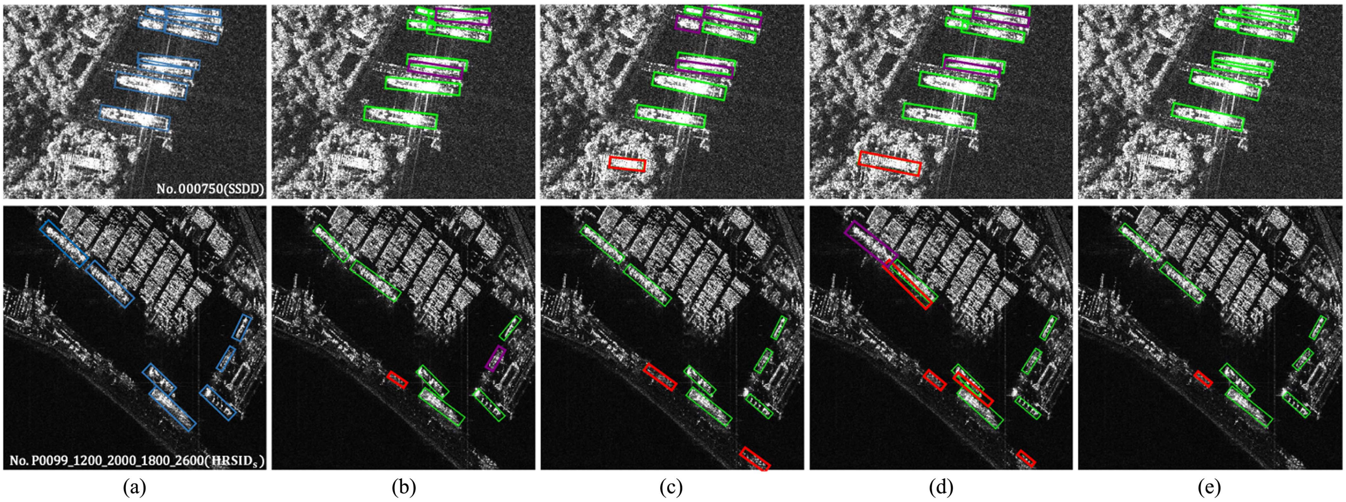


Fig. 8. Two detection examples of four detectors. (a) GT. (b) S^{2}ANet. (c) ReDet. (d) R^{3}Det. (e) Proposed. The blue, green, red, and purple represent GT, high-quality OBBs, false alarms, and missed ships.

TABLE V
EVALUATION ON DIFFERENT DETECTORS

Method	SSDD		HRSID$_s$	
	AP(%)	BEP	AP(%)	BEP
S^{2}ANet [64]	93.49	0.917	83.64	0.830
ReDet [65]	94.81	0.919	**84.73**	0.835
R^{3}Det [54]	93.86	0.930	82.73	0.827
R-RetinaNet+ [43]	94.66	0.940	-	-
Proposed	**96.47**	**0.949**	84.32	**0.840**

E. Comparison With State of the Art

We compare the proposed detector with several state-of-the-art OBB-based detectors, i.e., R^{3}Det [54], S^{2}ANet [64], R-RetinaNet+[43], ReDet [65]. S^{2}ANet and ReDet are retrained and tested based on MMDetection toolbox. For R-RetinaNet+, since we do not have access to the code, we directly compare with the results in the literature [43]. Therefore, there is no result of R-RetinaNet+ on HRSID$_s$ and only its result on SSDD is used in this article.

1) Overall Performance: According to Table V, on SSDD dataset, the proposed detector achieves the highest AP (96.47%), which is 1.66%, 1.81%, 2.61%, 2.98% higher than ReDet, R-RetinaNet+, R^{3}Det, S^{2}ANet, respectively. From Fig. 7(a), the curve of the proposed method is significantly better than other methods are significantly better than other methods, when the R is the same, the proposed method can achieve a higher P, and when the P is the same, a higher R can be achieved. On HRSID$_s$ dataset, the AP of the proposed detector ranks second, only 0.41% lower than that of ReDet [65]. Fig. 7(b) shows that this tiny gap is mainly attributed to that ReDet is slightly better than the proposed method in the low R high P and low P high R regions. However, in real application scenes, we tend to prefer the performance of the detector in the high P and high R region. Comparing Fig. 7(a) and (b), it can be seen that the performance of several detectors on HRSID$_s$ is not as good as on SSDD, mainly because HRSID$_s$ contains a large number of small ships (As shown in Fig. 1), the texture features of small ships in SAR images far inferior to large ships, and are easily confused with backgrounds so that small ships are easy to be missed. The performance gap of several detectors in HRSID$_s$ is not large, which shows that it is very difficult to ensure high R and high P for small ship detection. Fortunately, the proposed can improve in this regard, although the improvement is not quite significant, the improvement brought by the proposed is comprehensive. As shown in Fig. 7, the PR-curves of the proposed detector are consistently above the other detectors in high P and high R region, which indicates that the proposed one has better overall performance.

2) Balance Ability: The proposed detector exhibits an ability to balance R and P that is better than other methods, which can be derived from comparing the BEP of different methods in Table V. The proposed method achieves the highest BEP on both datasets, i.e., 0.949 on SSDD and 0.840 on HRSID$_s$. The BEP of several detectors on HRSID$_s$ is inferior to that on SSDD, and the BEP difference of different detectors on SSDD is larger than on HRSID$_s$, which once again shows the difficulty of detecting small ships, and some current studies are working on this problem. However, the proposed method does not specifically design for small ship detection, it still achieves performance improvements, which is enough to demonstrate its effectiveness. Besides, it can also be seen from Fig. 7 that the proposed detector can simultaneously achieve high R and high P, which indicates its balance ability is better.

3) In-Depth Analysis: In order to provide a more intuitive comparison and reveal the reasons behind the quantitative data, the detection results of two randomly selected test samples No.000750 from SSDD and No.P0099_1200_2000_1800_2600 from $HRSID_s$ are shown in Fig. 8. It shows that in inshore environment, the land buildings are similar to ship's appearance and ships are docked side by side have brought a great challenge to detection and postprocessing. All detectors produce false alarms, in which ReDet produces 3(1+2), R^3 Det produces 5(1+4), while S^2ANet and the proposed detector only produce 1(0+1). The false alarms are mainly due to the poor discriminative features extracted by network, resulting in some backgrounds being identified as ships by the classification subnet, but the proposed detector can extract better ship features to avoid it. In terms of missed ships, the detectors that missed ships all use NMS and the missed ships are mainly in the areas where ships are parked side by side. The proposed G-NMS is designed to consider this situation, so only the proposed detector does not miss ships. In addition, comparing Fig. 8(d) and (e), the proposed detector well solves the problems of false alarms and missed ships that exist in the original R^3Det. This result once again proves the advantages of the proposed detector in extracting discriminative representation features through the backbone network with MCCA and removing redundancy and false alarms by G-NMS. However, the proposed detector is not perfect, and false alarms may occur in the inshore scenes when backgrounds behave extremely similar to ships, as shown in Fig. 8(e). The generation of false alarms indicates that the background features extracted by the detector have a certain degree of similarity with ship features. Although the proposed method improves the discrimination of ship features well, there is still room for improvement.

IV. Conclusion

Focusing on the unsolved issues of SAR ship detection task in inshore and densely distributed scenarios, this article proposes a novel oriented SAR ship detector that achieves high *recall* and *precision* together. To enhance the ship feature representation extracted by the backbone network, we propose an MCCA module to reweight ship features so that important features are highlighted. In addition, G-NMS is proposed based on the geometric position relationship of ships, which effectively solves the problem of missed detection in scenes where ships are docked side by side and densely distributed. The detector with MCCA and G-NMS also achieves better performance. Extensive experiments on SSDD and $HRSID_s$ datasets validate the effectiveness and advantages of the proposed method.

Acknowledgment

The authors would like to thank the contributors to the SSDD and HRSID datasets for providing data support for this research. Meanwhile, the authors would like to thank all peer reviewers and associate editor for their constructive comments and suggestions that significantly improved this article.

References

[1] S. Brusch, S. Lehner, T. Fritz, M. Soccorsi, A. Soloviev, and B. van Schie, "Ship surveillance with TerraSAR-X," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1092–1103, Mar. 2011.

[2] W. An, C. Xie, and X. Yuan, "An improved iterative censoring scheme for CFAR ship detection with SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4585–4595, Aug. 2014.

[3] M. Kang, K. Ji, X. Leng, and Z. Lin, "Contextual region-based convolutional neural network with multilayer fusion for SAR ship detection," *Remote Sens.*, vol. 9, no. 8, 2017, Art. no. 860.

[4] D. Li, Q. Liang, H. Liu, Q. Liu, H. Liu, and G. Liao, "A novel multidimensional domain deep learning network for SAR ship detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, Mar. 2021.

[5] H. Lang, Y. Xi, and X. Zhang, "Ship detection in high-resolution SAR images by clustering spatially enhanced pixel descriptor," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5407–5423, Aug. 2019.

[6] J. Ai et al., "Robust CFAR ship detector based on bilateral-trimmed-statistics of complex ocean scenes in SAR imagery: A closed-form solution," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 3, pp. 1872–1890, Jun. 2021.

[7] H. Lang, S. Wu, and Y. Xu, "Ship classification in SAR images improved by AIS knowledge transfer," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 439–443, Mar. 2018.

[8] F. Sharifzadeh, G. Akbarizadeh, and Y. S. Kavian, "Ship classification in SAR images using a new hybrid CNN–MLP classifier," *J. Indian Soc. Remote Sens.*, vol. 47, no. 4, pp. 551–562, 2019.

[9] Y. Xu and H. Lang, "Ship classification in SAR images with geometric transfer metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6799–6813, Aug. 2021.

[10] N. Aghaei, G. Akbarizadeh, and A. Kosarian, "GreyWolfLSM: An accurate oil spill detection method based on level set method from synthetic aperture radar imagery," *Eur. J. Remote Sens.*, vol. 55, pp. 181–198, 2022.

[11] X. Ma, J. Xu, P. Wu, and P. Kong, "Oil spill detection based on deep convolutional neural networks using polarimetric scattering information from sentinel-1 SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022.

[12] F. Ma, F. Zhang, D. Xiang, Q. Yin, and Y. Zhou, "Fast task-specific region merging for SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, Jan. 2022.

[13] F. Ma, F. Zhang, Q. Yin, D. Xiang, and Y. Zhou, "Fast SAR image segmentation with deep task-specific superpixel sampling and soft graph convolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, Sep. 2021.

[14] R. Shang et al., "SAR image segmentation based on constrained smoothing and hierarchical label correction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, May 2021.

[15] Z. Tirandaz, G. Akbarizadeh, and H. Kaabi, "Polsar image segmentation based on feature extraction and data compression using weighted neighborhood filter bank and hidden Markov random field-expectation maximization," *Measurement*, vol. 153, 2020, Art. no. 107432.

[16] U. Kanjir, H. Greidanus, and K. Oštir, "Vessel detection and classification from spaceborne optical images: A literature survey," *Remote Sensing Environ.*, vol. 207, pp. 1–26, 2018.

[17] G. Gao, "A parzen-window-kernel-based CFAR algorithm for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 557–561, May 2011.

[18] J. Ai, X. Qi, W. Yu, Y. Deng, F. Liu, and L. Shi, "A new CFAR ship detection algorithm based on 2-D joint log-normal distribution in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 806–810, Oct. 2010.

[19] X. Leng, K. Ji, K. Yang, and H. Zou, "A bilateral CFAR algorithm for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 7, pp. 1536–1540, Jul. 2015.

[20] G. Gao, K. Ouyang, Y. Luo, S. Liang, and S. Zhou, "Scheme of parameter estimation for generalized gamma distribution and its application to ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1812–1832, Mar. 2017.

[21] Z. Zou, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," 2019 *arXiv:1905.05055*.

[22] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019.

[23] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[24] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2961–2969.

[25] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6154–6162.

[26] W. Liu et al., "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.

[27] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.
[28] J. Redmon and A. Farhadi, “Yolo9000: Better, faster, stronger,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 7263–7271.
[29] J. Redmon and A. Farhadi, “YOLOv3: An incremental improvement,” 2018, *arXiv:1804.02767*.
[30] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal speed and accuracy of object detection,” 2020, *arXiv:2004.10934*.
[31] N. Davari, G. Akbarizadeh, and E. Mashhour, “Corona detection and power equipment classification based on GoogleNet-AlexNet: An accurate and intelligent defect detection model based on deep learning for power distribution lines,” *IEEE Trans. Power Del.*, vol. 37, no. 4, pp. 2766–2774, Aug. 2022.
[32] G. Zhang, S. Lu, and W. Zhang, “CAD-net: A context-aware detection network for objects in remote sensing imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10015–10024, Dec. 2019.
[33] X. Zhou et al., “EAST: An efficient and accurate scene text detector,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5551–5560.
[34] Y.-L. Chang, A. Anagaw, L. Chang, Y. C. Wang, C.-Y. Hsiao, and W.-H. Lee, “Ship detection based on YOLOv2 for SAR imagery,” *Remote Sens.*, vol. 11, no. 7, 2019, Art. no. 786.
[35] Z. Cui, Q. Li, Z. Cao, and N. Liu, “Dense attention pyramid networks for multi-scale ship detection in SAR images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8983–8997, Nov. 2019.
[36] Z. Deng, H. Sun, S. Zhou, and J. Zhao, “Learning deep ship detector in SAR images from scratch,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 4021–4039, Jun. 2019.
[37] Z. Lin, K. Ji, X. Leng, and G. Kuang, “Squeeze and excitation rank faster R-CNN for ship detection in SAR images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 751–755, May 2019.
[38] Y. Zhao, L. Zhao, B. Xiong, and G. Kuang, “Attention receptive pyramid network for ship detection in SAR images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2738–2756, May 2020.
[39] X. Yang, X. Zhang, N. Wang, and X. Gao, “A robust one-stage detector for multiscale ship detection with complex background in massive SAR images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, Nov. 2021.
[40] W. Bao, M. Huang, Y. Zhang, Y. Xu, X. Liu, and X. Xiang, “Boosting ship detection in SAR images with complementary pretraining techniques,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8941–8954, Sep. 2021.
[41] T. Zhang et al., “Balance scene learning mechanism for offshore and inshore ship detection in SAR images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Nov. 2020.
[42] R. Yang, G. Wang, Z. Pan, H. Lu, H. Zhang, and X. Jia, “A novel false alarm suppression method for CNN-based SAR ship detector,” *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 8, pp. 1401–1405, Aug. 2021.
[43] R. Yang, Z. Pan, X. Jia, L. Zhang, and Y. Deng, “A novel CNN-based detector for ship detection based on rotatable bounding box in SAR images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1938–1958, Jan. 2021.
[44] Z. Sun, X. Leng, Y. Lei, B. Xiong, K. Ji, and G. Kuang, “BiFA-YOLO: A novel YOLO-based method for arbitrary-oriented ship detection in high-resolution SAR images,” *Remote Sens.*, vol. 13, no. 21, 2021, Art. no. 4209.
[45] Y. He, F. Gao, J. Wang, A. Hussain, E. Yang, and H. Zhou, “Learning polar encodings for arbitrary-oriented ship detection in SAR images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3846–3859, Mar. 2021.
[46] Z. Cui, X. Wang, N. Liu, Z. Cao, and J. Yang, “Ship detection in large-scale SAR images via spatial shuffle-group enhance attention,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 379–391, Jan. 2021.
[47] J. Fu, X. Sun, Z. Wang, and K. Fu, “An anchor-free method based on feature balancing and refinement network for multiscale ship detection in SAR images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1331–1344, Feb. 2021.
[48] Z. Sun et al., “An anchor-free detection method for ship targets in high-resolution SAR images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7799–7816, Jul. 2021.
[49] J. Zhang, M. Xing, G.-C. Sun, and N. Li, “Oriented gaussian function-based box boundary-aware vectors for oriented ship detection in multiresolution SAR imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, Jul. 2021.
[50] K. Fu, J. Fu, Z. Wang, and X. Sun, “Scattering-keypoint-guided network for oriented ship detection in high-resolution and large-scale SAR images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 11162–11178, Sep. 2021.
[51] A. Neubeck and L. Van Gool, “Efficient non-maximum suppression,” in *Proc. 18th Int. Conf. Pattern Recognit.*, 2006, pp. 850–855.
[52] J. Li, C. Qu, and J. Shao, “Ship detection in SAR images based on an improved faster R-CNN,” in *Proc. SAR Big Data Era: Models Methods Appl.*, 2017, pp. 1–6.
[53] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, “HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation,” *IEEE Access*, vol. 8, pp. 120234–120254, 2020.
[54] X. Yang, J. Yan, Z. Feng, and T. He, “R3det: Refined single-stage detector with feature refinement for rotating object,” in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 4, pp. 3163–3171, 2021.
[55] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
[56] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, “ECA-net: Efficient channel attention for deep convolutional neural networks,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11531–11539.
[57] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.
[58] W. Qian, X. Yang, S. Peng, Y. Guo, and J. Yan, “Learning modulated loss for rotated object detection,” 2019, *arXiv:1911.08299*.
[59] R. Girshick, “Fast R-CNN,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1440–1448.
[60] K. Chen et al., “MMDetection: Open MMLab detection toolbox and benchmark,” 2019, *arXiv:1906.07155*.
[61] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis, “Soft-NMS–improving object detection with one line of code,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 5561–5569.
[62] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, “Distance-IoU loss: Faster and better learning for bounding box regression,” in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 12993–13000.
[63] R. Solovyev, W. Wang, and T. Gabruseva, “Weighted boxes fusion: Ensembling boxes from different object detection models,” *Image Vis. Comput.*, vol. 107, 2021, Art. no. 104117.
[64] J. Han, J. Ding, J. Li, and G.-S. Xia, “Align deep features for oriented object detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, Mar. 2021.
[65] J. Han, J. Ding, N. Xue, and G.-S. Xia, “ReDet: A rotation-equivariant detector for aerial object detection,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 2786–2795.

Chunnan Li received the B.E. degree in electronic science and technology from the Beijing University of Chemical Technology (BUCT), Beijing, China, in 2020. He is currently working the M.S. degree in physics with the College of Mathematics and Physics, BUCT.

His research interest is algorithms for ship detection in SAR remote sensing imagery using computer vision.

Haitao Lang (Member, IEEE) received the B.E. and M.S. degrees in optical engineering from Ocean University of China, Qingdao, China, in 2000 and 2003, respectively, the Ph.D. degree in optical engineering from the Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai, China, in 2006.

He is currently a Professor with the College of Mathematics and Physics, Beijing University of Chemical Technology, Beijing, China. He has authored prolifically in some prestigious refereed journals and conference proceedings, such as IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, Conference on Computer Vision and Pattern Recognition (CVPR), Asian Conference on Computer Vision (ACCV), International Geoscience and Remote Sensing Symposium (IGARSS), and SPIE Remote Sensing. His research interests include machine learning, pattern recognition, and optical engineering with a focus on developing advanced image analysis and interpretation techniques for maritime remote sensing applications.